View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Huddersfield Repository

University of Huddersfield Repository

Leichsenring, Falk and Rabung, Sven

The Role of Efficacy vs. Effectiveness Research in Evaluating Psychotherapy

Original Citation

Leichsenring, Falk and Rabung, Sven (2007) The Role of Efficacy vs. Effectiveness Research in Evaluating Psychotherapy. Mental Health and Learning Disabilities Research and Practice, 4 (2). pp. 125-143. ISSN 1743-6885

This version is available at http://eprints.hud.ac.uk/12381/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/

Mental Health and Learning Disabilities Research and Practice, 2007, 125 -143

The Role of Efficacy vs. Effectiveness Research in Evaluating Psychotherapy

Falk Leichsenring[1] & Sven Rabung[2]

[1] Department of Psychosomatics and Psychotherapy, University of Giessen, Germany

[2] Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

The Role of Efficacy vs. Effectiveness Research in Evaluating Psychotherapy

Falk Leichsenring[1] & Sven Rabung

Abstract

A sound base of evidence for different kinds of psychotherapy is a fundamental prerequisite for adequate access to treatment. The present article addresses the question of what kind of evidence is required to demonstrate that a specific method of psychotherapy works. Referring to recent conceptualisations of the logical structure of scientific theories, the authors argue that randomised controlled trials (RCTs) and effectiveness studies refer to different domains of intended applications. In RCTs the efficacy of a treatment under controlled experimental conditions is tested, whereas in effectiveness studies the outcome of a treatment under the conditions of clinical practice is examined. Accounting for the different domains of intended applications has several important implications. These implications refer to methodological questions of research (e.g. internal and external validity, study design type), but also the possibility of transferring results from experimental to field settings, and the criteria for evaluating the evidence-base of treatments. They also address the important question of the degree to which the presently available forms of psychotherapy can be regarded as evidence-based. The considerations presented in this article primarily refer to psychotherapy; however, they are also relevant to the evidence-based treatment approach in general.

Key words: Randomised controlled studies, effectiveness studies, philosophy of science, psychotherapy research, levels of evidence, empirically supported treatments.

With the recent emphasis on evidence-based medicine (EBM) it is not surprising that a corresponding approach has reached the field of psychotherapy. A prominent example is the empirically supported treatment (EST) approach of the Task Force of Division 12 (Clinical Psychology) of the American Psychological Association (APA) (Chambless & Hollon 1998; Chambless & Ollendick 2001; Task Force on Promotion and Dissemination of Psychological Procedures 1995). The EST approach has been much discussed with regard to whether it is appropriate for psychotherapy (Beutler 1998; Fonagy 1999; Leichsenring 2004; Persons & Silberschatz 1998; Roth & Parry 1997; Rothwell 2005; Seligman 1995; Westen, Novotny, & Thompson-Brenner 2004). The present paper addresses this question in a novel way. The evidence available for different psychotherapeutic methods is highly relevant to the question of which treatments are recommended for which patients.

Levels of Evidence

Several methods have been proposed for grading the available evidence of both

medical and psychotherapeutic treatments (Canadian Task Force on the Periodic Health Examination 1979; Chambless and Hollon 1998; Clark & Oxman 2003; Cook et al. 1995; Guyatt et al. 1995; Nathan & Gorman 2002; National Institute of Clinical Excellence 2002). Despite some differences, all available proposals regard RCTs (efficacy studies) as the "gold standard" for demonstrating that a treatment is efficacious. For example, according to the APA's criteria for an EST, at least two RCTs (or controlled single-case experiments or equivalent time-samples designs) from independent research groups are required in which a therapy group is significantly superior to a no-treatment group, placebo group, an alternative treatment or equivalent to an already established therapy (Chambless & Hollon 1998; Chambless & Ollendick 2001; Task Force on Promotion and Dissemination of Psychological Procedures 1995). Furthermore, a treatment manual must be used and a specific mental disorder must be studied. For the designation of being efficacious and specific, a treatment must have been shown to be superior to pill or psychological placebos or to an alternative bona fide treatment in at least two independent RCTs.

Randomised Controlled Studies and Levels of Evidence

RCTs are conducted under controlled experimental conditions; thus they allow experimenters to control for variables that may systematically influence the outcome independent of the treatment. The defining feature of an RCT is the random assignment of subjects to different conditions of treatment (Shadish, Cook, & Campbell 2002). Randomisation is regarded as indispensable in order to ensure that differences between subjects are equally distributed. The goal of randomisation is to attribute the observed effects exclusively to the applied therapy. Thus, randomisation is used to ensure the internal validity of a study, i.e. the validity of inference about whether the observed covariation between treatment and outcome reflects a causal relationship from treatment to outcome (Shadish et al. 2002). External validity refers to the question if the observed relationship between treatment and outcome specified in the hypothesis under study holds for different patients, therapists, setting and measures. Within efficacy research the most stringent test is achieved by comparison with rival treatments, thus controlling for specific and non-specific therapeutic factors (Chambless & Hollon 1998, p. 8; Gabbard, Gunderson & Fonagy 2002). Furthermore, such comparisons provide explicit information regarding the relative benefits of competing treatments. Treatments that are found to be superior to rival treatments are more highly valued. Gabbard et al. discuss different types of RCTs that provide different levels of evidence (Gabbard, Gunderson, & Fonagy 2002). The authors regard RCTs in which a treatment is compared to a psychological placebo as the second most rigorous variant within RCTs. However, in our view comparisons with treatments as usual (TAU) can provide more stringent tests than placebo controlled studies, because they control for both common factors (e.g. attention) and treatment effects of TAU. However, a frequent problem in TAU-controlled studies is that TAU is poorly defined and differs from one study to another. In one study, for example, TAU may be routine outpatient psychotherapy in clinical practice, whereas in another study it may be a pure psychopharmacological treatment, and in a third study it may include counselling or other forms of health care.

The fourth most rigorous form of RCTs uses waiting list controls. However, in this type of study, it is not clear if the observed effect in the treatment group is to be attributed to specific or non-specific therapeutic factors (Gabbard et al. 2002). According to Gabbard et al. (2002), the next level of evidence is provided by prospective pre-post studies, followed by case series and finally by case reports.

Critique of the RCT Approach

The EBM and EST approaches, with their emphasis on RCTs, follow the methodology of pharmacological research. It has been critically discussed whether this methodology is adequate for psychotherapy research (Beutler 1998; Fonagy 1999; Leichsenring 2004; Persons & Silberschatz 1998; Roth & Parry 1997; Rothwell 2005; Seligman 1995; Westen, Novotny & Thompson-Brenner 2004). The main arguments can be summarized as follows: (1) The defining features of RCTs such as randomisation, use of treatment manuals, focus on specific mental disorders and frequent exclusion of patients with a poor prognosis (e.g. multimorbid patients) raise the question as to whether the results of RCTs are sufficiently representative of clinical practice. (2) Thus, if a method of psychotherapy has been shown to work under the controlled conditions of an RCT, this does not necessarily imply that it equally works under the conditions of clinical practice (Leichsenring, 2004). (3) The EST approach puts the emphasis on disorders and on symptoms (Blatt 1995). As Henry (1998, p.129) put it: "EVTs [Empirically Validated Treatments] place the emphasis on the disorder ... and not on the individual... who seeks our services." (4) At present, treatment manuals do not exist for multimorbid patients as they usually occur in clinical practice. The available treatment manuals refer to the treatment of isolated mental disorders. It is widely unknown how the treatment of one specific mental disorder (e.g. depressive disorder) must be modified according to coexisting disorders (e.g. social phobia or narcissistic personality disorder). (5) The methodology of RCT with its use of treatment manuals and randomised control conditions is hardly applicable to long-term psychotherapy lasting several years (Seligman 1995; Wallerstein 1999). (6) The APA criterion of placebo control groups can be regarded as questionable (Leichsenring 2004). The concept of placebo controls (nonspecific or common factor controls) in psychotherapy research is so conceptually flawed that Lambert and Bergin (1994, p. 152) pleaded, over 10 years ago, to give up placebo controls in psychotherapy research. Placebo effects in psychotherapy are, in the end, psychotherapeutic effects. (7) From a methodological perspective, in many RCTs randomisation is only a formal criterion, and the intended control over confounding variables is probably not achieved. In order to control for differences between subjects by randomisation, a sufficient number of subjects is necessary (Hsu 1989). In many RCTs the sample size is not large enough to achieve equivalent groups by randomisation (Hsu 1989; Leichsenring & Rabung 2006). For example, in the meta-analysis by Gloaguen, Cottraux, Cucherat, and Blackburn (1998) of the efficacy of cognitive therapy in depression, only 38% of the groups had a sample size of at least 20 patients and only 13% of samples had a size of at least 40 patients per group, thus fulfilling the criteria for an effective randomisation as formulated by Hsu (1989). Furthermore, patients frequently

drop out from control groups, additionally restricting the comparability and representativeness of the control groups to the treatment group even in RCTs.

Effectiveness Studies

Contrary to RCTs, effectiveness studies are carried out under the conditions of clinical practice. They are highly representative for clinical practice (Shadish, Matt, Navarro, & Phillips 2000): patients with complex (i.e. highly comorbid) disorders (Guthrie 2000), as they usually occur in clinical practice, are treated. Therapists apply exactly those methods of psychotherapy that they are accustomed to and experienced in. Patients are referred to the respective treatments in the standard way of clinical practice, including their own preferences and decisions regarding a specific kind of therapy or for a specific psychotherapist. The duration of the treatment is determined by their clinical requirements (Seligman 1995). Thus, effectiveness studies provide evidence of the results of a treatment under the conditions of clinical practice. For this reason, Barkham et al. (2001) and Lucock et al (2003) used the term practice-based evidence to characterize data from the effectiveness research approach as contrasted to the evidence-based treatment approach (efficacy data). These authors have described a systematic approach to generate practice-based evidence for routine therapy services. The National Institute of Mental Health in the USA (NIMH) has specifically called for more effectiveness research (Krupnick et al. 1996). The U.K. Department of Health (1999) stressed the need for evaluating psychotherapy services in routine service conditions and the need to compare outcome data from routine clinical practice with those obtained from RCTs (U.K. Department of Health, 1996).

Limitations of Effectiveness Studies

According to the considerations made above, the strength of effectiveness studies is their clinical representativeness as they are carried out under conditions of routine practice. As a consequence, however, effectiveness studies cannot control for factors affecting the outcome to the same extent as RCTs (internal validity). Thus, the main argument against effectiveness studies for demonstrating whether a treatment works refers to possible threats to internal validity, i.e., to the reduced ability to control factors influencing the outcome independent of therapy. This is likely to be the reason why effectiveness studies are not accepted, for example, by the American Psychological Association as a method for demonstrating that a therapy works. However, there is evidence that effectiveness studies do not seem to overestimate effect sizes compared to RCTs. This evidence refers not only to psychotherapy (Shadish et al. 2000), but also to the broader field of EBM (Benson & Hartz 2000; Concato, Shah, & Horwitz 2000). For the field of EBM, Benson and Hartz (2000) and Concato, Shah, and Horwitz (2000) compared the effects of the same treatment applied to a specific condition found in effectiveness studies (”observational studies”) with those of RCTs. They did not find systematic differences. They conclude that there is little evidence that effect sizes of well-designed effectiveness studies are larger than or qualitatively different from those found in RCTs. Shadish et al. (2000) rated the clinical representativeness of selected psychotherapy outcome studies. Non

randomised studies were rated as significantly more clinically representative. However, the authors did not find a correlation between clinical representativeness and effect size. Thus, it can be concluded that non-randomised studies did not overestimate effect sizes compared to those obtained in RCTs.[2] After all, these findings parallel those reported by Benson and Hartz (200) and Concato et al (2000).

Philosophy of Science and an Alternative Perspective

Taking more recent developments in the philosophy of science into account, it has recently been argued that EBM and EST are implicitly based on an outdated conception of the logical structure of scientific theories (Leichsenring 2004), that being the statement view as the standard conception of scientific theories (Hempel 1970). The statement view assumes universal validity or application of a theory. Regarding RCTs as the "gold standard" for psychotherapy outcome research is implicitly based on the statement view. If, and only if, a therapy has been proven to work in a controlled research setting (laboratory) can it then be applied to clinical practice (the field), based on the assumption that it works equally well in the field.

The statement view of scientific theories can be contrasted with the structuralist conception of scientific theories (Sneed 1971; Stegmüller 1979; Westmeyer 1982, 1989). In the structuralist conception of scientific theories the domain of intended applications is regarded as an integral component of a theory and the hypotheses derived from it. According to this conception a theory consists of a theory-core K and a set I of intended applications of K: "The inclusion of this latter set I into the definition of a theory-element is characteristic of the structuralist approach. It is a fundamental tenet of structuralism, that a theory is not universally applicable, but only to a certain set, range, or domain of intended applications" (Westmeyer 1989, p. 4). In order to apply a theory, it is necessary to extend the theory-core. This is achieved by stating hypotheses and special conditions that are valid only for the intended applications. Therefore, from a structuralist viewpoint, there are no context-free hypotheses; hypotheses always refer to specific contexts or intended applications. In RCTs, hypotheses about the efficacy of treatments under controlled experimental (idealized) conditions are tested; the selection of patients, therapists, treatments, and outcome measures takes place within a research project. Thus, in RCTs, laboratory-based hypotheses and modifications of real-life therapies are tested. For the latter, the term "laboratory therapies" has been proposed (Leichsenring 2004). In effectiveness studies hypotheses referring to the conditions of clinical practice are tested (field hypotheses and "field therapies"). Thus, it depends on the intended applications whether an efficacy or effectiveness studies is to be carried out. If the hypothesis under study refers to laboratory contexts, an RCT is required; if the hypothesis refers to natural conditions, an effectiveness study is required.

[2] Some highly clinically representative non randomised studies rather underestimated the effect sizes. The authors attribute the smaller effect sizes of these studies to self-selection of more distressed patients to the treatment group and of the less distressed patients to the control group.

Thus, from a structuralist viewpoint, controlled and effectiveness studies do not fundamentally differ in principle concerning their external validity, that is the degree to which the patients, therapists, treatments, settings and outcome measures are representative for the conditions specified in the hypothesis under study, in other words, for the domain of intended applications. Furthermore, there is no difference between RCTs and effectiveness studies concerning their internal validity. Extensions of the theory-core and intended applications in reference to laboratory conditions will lead to simpler hypotheses and to more rigorous special conditions. In contrast, extensions of the theory-core and intended applications in reference to field conditions will have to compensate for reduced experimental control with more liberal special conditions and correspondingly extended hypotheses. A difference in internal validity does not arise (for a more detailed discussion see Westmeyer 1982). Thus, from a structuralist view, RCTs and effectiveness studies do not fundamentally differ concerning their internal or external validity, and RCTs do not necessarily provide higher-level evidence than effectiveness studies.

As efficacy studies provide evidence from an intended application different from clinical practice, empirical evidence from RCTs cannot be directly transferred to the conditions of the field. If a method of psychotherapy has been shown to work under laboratory conditions (efficacy), this does not necessarily imply that it equally works under natural conditions (effectiveness). There is no justification for an inductive generalization from experimental to non-experimental conditions of everyday reality (Bredenkamp 1980; Leichsenring 1985). The effectiveness of a treatment under natural conditions can be demonstrated only by effectiveness studies. For this reason, the RCTs listed by the APA (Chambless & Hollon 1998; Chambless & Ollendick 2001) as empirical support for specific psychotherapeutic methods show only that these methods work under experimental (laboratory) conditions. That they work equally well in the field has not yet been demonstrated. One of the main reasons for the gap between experimental (laboratory) conditions and clinical practice is that psychotherapy is not a drug that works equally under different conditions. Difficult-to-quantify factors in the therapist-patient match may influence outcome. Thus, it is questionable whether the methodology of pharmacological research is adequate for psychotherapy research of mental disorders, at least when the effectiveness of a treatment in clinical practice is to be studied. After all, RCTs serve only a limited function (Roth & Parry 1997).

Levels of Evidence: an Alternative Perspective

Taking into account the different intended applications associated with efficacy and effectiveness studies, the schemes of levels of evidence which regard RCTs as the gold standard (Canadian Task Force on the Periodic Health Examination 1979; Chambless & Ollendick 2001; Cook et al. 1995; Guyatt et al. 1995; Nathan & Gorman, 2002) refer to treatments under laboratory conditions (efficacy studies). Thus, they cannot be applied to the question of whether a therapy works under conditions of routine practice. For effectiveness studies, which by definition cannot use randomisation, levels of evidence must

be defined by criteria different from those of efficacy studies.[3] For this reason, it has been proposed that the criteria and levels of evidence for RCTs should be separated from those for effectiveness studies (Leichsenring 2004).

Levels of Evidence of Effectiveness Studies

For the empirical support of treatments applied in field settings, another parallel scheme is necessary to describe the levels of evidence in effectiveness studies. Such a scheme has to take into account the quality of effectiveness studies. The criteria defining the quality of effectiveness studies cannot be identical to those of RCTs, although they may overlap. High-level evidence for the effectiveness of a treatment is provided by high-level effectiveness studies. By what criteria can high-level effectiveness studies be defined?

Shadish et al. (2002) have described experimental and quasi-experimental designs for generalized causal inference. According to Shadish et al. (2002), a causal inference from a quasi-experimental study must meet three basic requirements: cause must precede effect, cause must co-vary with effect, and alternative explanations of the effect must be implausible. As quasi-experimental studies do not use random assignment, they have to use other principles to show that alternative explanations of the effect are implausible. These principles include:

1. The identification and study of plausible threats to internal validity,

2. The use of additional design elements (e.g., observation at more pre-test time points, additional comparison groups) or of statistical controls, and

3. Coherent pattern matching, that is, prediction of complex patterns of results (e.g., non-equivalent dependent variables or interactions).

Shadish et al. (2002) discuss several measures of control for different types of quasi-experimental designs, i.e. for quasi-experimental designs without control groups or pre-tests or for quasi-experimental designs with control groups and pre-tests. Due to space limitations, these measures cannot be described here in detail. A detailed presentation is given by Shadish et al. (2002). The prediction of differential effects or interactions as a function of different methods of therapy, different outcome measures or moderator variables may serve as an example of coherent pattern matching (for the definition of moderator variables, see Kendall, Holmbeck & Verduin 2004). For example, the results of cognitive-behavioural therapy of social phobia are modified by several moderator variables such as the severity of symptoms (Otto et al. 2000), type of social phobia (isolated or general, Hope et al. 1995), comorbid

[3] There is only one exception: If a waiting list group is used as a comparison condition, randomisation can be used in an effectiveness study without affecting the service delivery too much. Patients as they are usually treated in clinical practice can be randomly assigned to the treatment or to the waitlist group. I would like to thank one of the anonymous reviewers for calling our attention to point. - However, for ethical and practical reasons the use of waiting list comparison groups is only justifiable and applicable for short-term treatments.

depression (Erwin et al. 2002), and expectations of therapy (Chambless, Tran, & Glass 1997). When included in a study design, these variables can be examined with regard to their effects on therapy outcome. The more that the predicted differential effects occur, the less probable it is that the changes observed can be attributed to factors other than the method of therapy applied. Furthermore, according to a recent proposal 'change norms' can be used as additional design elements (Leichsenring & Rabung 2006). Furthermore process research can contribute to a more stringent conclusion that an observed effect is associated with the interventions applied. In an open study of psychodynamic therapy for depression, Hilsenroth et al. (2003) showed that the observed effects were associated with the defining features of psychodynamic therapy but not with the defining features of cognitive-behavioural therapy.

Taking these refined methodological issues into account a proposal has been made to grade the levels of evidence of effectiveness studies (Leichsenring 2004). According to this proposal, a high-level effectiveness study is a prospective quasi-experimental study of high clinical representativeness, characterized by non-random comparison groups, the matching or stratifying of groups, clear descriptions of treatments, patients and their selection, the use of reliable and valid diagnostic procedures and outcome measures, the use of additional design elements, coherent pattern matching, reporting of drop outs, pre- and post-assessments, follow-up studies, and the reporting of relevant statistical data. Clinical representativeness is achieved by the selection of patients, therapists, and treatments that are typical for clinical practice (Wells 1999; Shadish et al. 2000). Plausible threats to internal validity are controlled for by the use of additional design elements (e.g., observation at more pretest time points, additional comparison groups), statistical controls, or coherent pattern matching, that is, prediction of complex patterns of results (e.g., non-equivalent dependent variables or interactions). According to this definition, the gold standard of effectiveness studies (effectiveness studies) is a prospective quasi-experimental study of high clinical representativeness that fulfills all or at least most of the aforementioned criteria. Lower-level effectiveness studies differ from high-level studies in one or more of these aspects (Leichsenring 2004). In order to judge the effectiveness of a method of therapy (in a specific disorder) in clinical practice, the existing studies have to be rated with regard to their levels of evidence according to criteria that remain to be defined. Furthermore, definitions must be given that are similar to those of the APA guidelines concerning the number of studies regarded as necessary. For a treatment to be judged as "effective" in clinical practice, at least two independent level I studies may be regarded as necessary. To be judged as "probably effective," one level I study may be regarded as necessary.

Complementary Relationship of Efficacy and Effectiveness Studies

As discussed above, efficacy and effectiveness studies address different research questions: RCTs examine the efficacy of a treatment under controlled experimental conditions, whereas effectiveness studies address the effectiveness under clinical practice conditions. As a consequence, the relationship between RCTs and effectiveness studies is not competitive;

rather, it is complementary. From this perspective, a distinction between empirically supported therapies (EST) and RCT methodology is required (Leichsenring 2004; Westen, Novotny, & Thompson-Brenner 2004).

Discussion

The present article addresses the discussion of efficacy vs. effectiveness in psychotherapy outcome research. Considerations from the view of the philosophy of science have shown that methodological questions cannot be answered in isolation from content, i.e. from the question of research. Buchkremer and Klingberg (2001) proposed that the testing of psychotherapy should be conceptualised as analogous to the testing of pharmacological treatments. As noted above, the essential argument against such a model is that psychotherapy is not a drug that works equally under different conditions, i.e. in the laboratory of a research project and in the field of clinical practice. The results of social psychology experiments alone (e.g. Rosenthal 1981) speak against such a model. Contrary to pharmacological research, interpersonal factors are not only the common, but also the specific curative factors (e.g., Luborsky 1984; Lambert & Bergin 1994; Orlinsky, Grawe, & Parks 1994). For this reason, psychotherapy outcome research needs a model of its own that is appropriate to its subject. Making RCTs an absolute is a result of the abuse of the drug metaphor in psychotherapy research (Stiles & Shapiro 1989).

The presently available proposals made to define evidence-based psychotherapeutic treatments refer to the treatment of a specific mental disorder (e.g. panic disorder). However, results of epidemiological studies have shown that most patients do not suffer from an isolated mental disorder (Kessler et al. 1994; Kessler, Chiu, Demler, & Walters 2005). Rates of comorbidity are typically high, and most patients are multimorbid. However, as mentioned above, the treatment manuals presently available describe the treatment of a specific isolated mental disorder. They do not answer the question of how to treat, for example, a major depressive disorder associated with a post-traumatic stress disorder and a narcissistic personality disorder. It is widely unknown how effective the empirically supported treatments for depressive disorders, as listed for example by Chambless and Hollon (1998) or Nathan and Gorman (2002), are in the treatment of highly comorbid patients. Furthermore, the emphasis on highly structured psychotherapeutic methods tailored to the treatment of a specific mental disorder corresponds to a conceptualisation of psychotherapy-like cognitive-behavioural therapy rather than to other forms of psychotherapy, e.g. psychodynamic therapy. Thus, it is no surprise that the EST approach was initiated by proponents of CBT (Chambless and Hollon 1998, Chambless and Ollendick 2001). Furthermore, a highly structured conceptualisation of psychotherapy seems to correspond to specific personality traits that were found in CBT therapists rather than in psychodynamic therapists (Topolinski & Hertel, 2007).

Certainly, the differentiation of randomised controlled (laboratory) studies and effectiveness studies, on which this article has focused, is not a new idea (e.g. Seligman 1995; Wells 1999). However, what is new is the pursuit of this differentiation with regard to the implications that arise from a structuralistic

view of theories: RCTs and effectiveness studies serve different purposes and answer different questions of research. Thus, RCTs are required, for example, if a newly developed method of psychotherapy is to be tested with regard to specific therapeutic effects. This is especially true if alternative treatments are available. For example, Crits-Christoph et al. (1999) tested a specialized form of Luborsky´s psychodynamic therapy for the treatment of cocaine dependence by comparing it to cognitive-behavioural therapy and drug counselling. RCTs are also required for the study of isolated elements of therapy (dismantling strategies, e.g. Borcovec 1993; Jacobson et al. 1996; Ahn & Wampold 2001). Before a new method of psychotherapy is applied in clinical practice, it is necessary for both financial and ethical reasons to test its efficacy in a preliminary step under controlled conditions. On the other hand, if the effectiveness of a treatment in the field is to be tested, effectiveness studies of high methodological quality are required. On the other hand, if a treatment has been shown to work under the conditions of clinical practice, it does not make sense to test its efficacy under experimental conditions.

We hope that the differentiation of empirical evidence into laboratory vs. field types of evidence and, respectively, into laboratory therapies vs. field therapies will be incorporated into psychotherapy research and, thus, in the conceptualisation of evidence-based psychological therapies. This differentiation implies that for many psychotherapeutic methods (e.g. Chambless & Ollendick 2001), there is evidence only that they work under laboratory conditions. It is widely unknown how effective these methods are in the field of psychotherapeutic practice. It is only in recent years that researchers have become aware of this problem. Shadish et al. (2000), for example, studied the influence of the degree of clinical representativeness of studies on the outcome of therapy.

From the differentiation into laboratory and field types of evidence, a new research agenda for effectiveness studies can be derived which is analogous to that of efficacy studies (Leichsenring 2004). This research addresses how effective psychotherapeutic methods in specific, though comorbid disorders (e.g. depression, anxiety disorders, somatoform disorders, highly comorbid disorders) are in clinical practice. There are at least two different strategies: In the first, the effectiveness of therapies that are already applied in the field is evaluated (e.g. Seligman 1995); in the second, psychotherapeutic methods that had been tested in RCTs are applied and, if necessary, modified in clinical practice (e.g. Hahlweg, Fiegenbaum, Frank, Schroeder, & von Witzleben 2001). At present, only a few studies of this type for specific disorders exist, e.g. for panic disorders (Wade, Treat, & Stuart 1998; Hahlweg et al. 2001), depression (Peterson & Halstead 1998; Persons, Bostrom & Bertagnolli 1999; Organista, Munoz & Gonzales 1994), bulimia nervosa (Tuschen-Caffier, Pook & Frank 2001) or for externalising disorders in children and adolescents (Tynan, Schumann & Lampert 1999). Initial results show, that in routine clinical practice patients do not profit from specific methods of therapy to the same extent as was reported from RCTs, that therapies are carried out for a longer time, or that additional elements of therapy are added, e.g., psychopharmacological therapy (Chambles & Ollendick 2001, p. 711). In other words, the laboratory forms of therapy are not purely applied in clinical

practice: modified versions are used. These results show that data from experimental conditions cannot directly be transferred to clinical practice.

The present article explicitly calls for a definition of the criteria that determine the level of quality of an effectiveness study. By definition, effectiveness studies cannot use randomisation with the exception of waiting-list comparison conditions. They have to apply other strategies to ensure internal validity. According to the extent to which internal, external, and other aspects of validity are ensured, different effectiveness studies may differ concerning their levels of evidence. Whereas external validity of effectiveness studies concerning the treated patients can be ensured relatively easily (e.g. by comparison with epidemiological data), this is more difficult with regard to therapists. For this purpose, data of relevant features of therapists are required (Wells 1999). Internal validity of both effectiveness studies and RCTs can be reduced by several factors, one being small sample size. With insufficient sample size, small differences between (nonrandomised) groups in effectiveness studies are not detected with sufficient power (Wells 1999). In RCTs it is questionable whether randomisation leads to equivalent groups if the sample size is insufficient (Hsu 1989). Dropouts can impair the internal validity in both types of studies. Here, intent-to-treat analyses are required (Wells 1999). Moreover, there is empirical evidence that randomisation is often incorrectly carried out, and that non-random manipulations of comparison groups are made (Schulz, Chalmers, Grimes & Altman 1994). The results presented by Shadish at al. (2000) are relevant with regard to the appropriateness of effectiveness studies as methods for testing if a treatment works. Shadish at al. (2000) did not find a significant correlation between the degree of clinical representativeness (e.g., RCTs vs. effectiveness studies) and the size of the effects reported in studies of psychotherapy. Thus, the conclusion can be drawn that (high-level) effectiveness studies do not systematically overestimate the effects of psychotherapy.

The proposal for levels of evidence made in this article can be used in both the judgment of existing effectiveness studies and planning of new studies. This proposal is intended to stimulate scientific discussion.

Finally, it should be noted that the differentiation between pure efficacy studies and pure effectiveness studies is somewhat arbitrary. The distinction between these two types of study designs can be more blurred than described in this paper so far (Guthrie. 2000). In RCTs, for example, elements of routine clinical practice can be implemented (e.g. treatments as they are usually applied in clinical practice or patients as they usually treated in clinical practice). Also in effectiveness studies elements of RCTs can be used, for example patients with specific mental disorders can be included. However, these patients usually show not only one, but multiple mental disorders. If randomisation is used in a study carried out under the conditions of clinical practice, the representativeness for clinical practice is reduced, because patients are not referred to the treatments by the usual ways of clinical practice including their own preferences for a specific type of treatment or for an individual therapist – the one exception was described above. Thus, a continuum may be more appropriate than a dichotomous distinction between experimental and effectiveness studies with “pure” RCTs marking the one pole and “pure”

effectiveness studies marking the other pole. Studies including elements of the other study type lie in between. Consistent with this mixed model, Hilsenroth et al. (2003) applied a model that they called (p. 349) a “hybrid efficacy/effectiveness treatment research model”.

References

Ahn, H. & Wampold, B.E. 2001. Where oh where are the specific ingredients? A meta-analysis of component studies in counselling & psychotherapy. Journal of Counselling Psychology 48: 251-257.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K. & McGrath, G. 2000. Service profiling and outcomes benchmarking using the core-om: Towards practice-based evidence in the psychological therapies. Journal of Consulting & Clinical Psychology 69: 184-196

Benson, K., & Hartz, A.J. 2000. A comparison of observational studies and randomised, controlled trials. New England Journal of Medicine 342: 1878-1886.

Beutler, L.E. 1998. Identifying empirically supported treatments: What if we didn't? Journal of Consulting & Clinical Psychology 66: 113-120.

Blatt, S. 1995. Why the gap between psychotherapy research and clinical practice: A response to Barry Wolfe. Journal of Psychotherapy Integration 5: 73–76.

Borcovec, T. 1993. Between-group therapy outcome research: Design and methodology. In Onken, L.S., Blaine, J.D. & Boren, J.J. (Eds.). Behavioural Treatments for Drug Abuse and Dependence (pp. 249-290). Rockville, MD, National Institute on Drug abuse.

Bredenkamp, J. 1980. Theorie des Experiments [Theory of the experiment]. Darmstadt: Steinkopf.

Buchkremer, G. & Klingberg, S. 2001. Was ist wissenschaftlich fundierte psychotherapie? Zur diskussion um leitlinien für die psychotherapieforschung [What is scientifically based psychotherapy?]. Nervenarzt 72: 20-30.

Canadian Task Force on the Periodic Health Examination 1979. The Periodic Health Examination. Canadian Medical Association Journal 121: 1193-1254.

Chambless, D.L., Tran, G.Q. & Glass, C.R. 1997. Predictors of response to cognitive-behavioural group therapy for social phobia. Journal of Anxiety Disorders 11: 221-240.

Chambless, D.L. & Hollon, S.D. 1998. Defining empirically supported treatments. Journal of Consulting and Clinical Psychology 66: 7-18.

Chambless, D.L. & Ollendick, T.H. 2001. Empirically supported psychological interventions: Controversies and evidence. Annual Review of Psychology 52: 685-716.

Clarke, M. & Oxman, A.D. 2003. Cochrane Reviewers´s Handbook 4.1.6 (updated January 2003). In The Cochrane Library, Issue 1, 2003. Oxford, Update Software. Updated quarterly.

Concato, J., Shah, N. & Horwitz, R.I. 2000. Randomised, controlled trials, observational studies, and the hierarchy of research designs. New England Journal of Medicine 342: 1887-1892.

Cook, D.J., Guyatt, G.H. Laupacis, A., Sacket, D.L., & Goldberg, R.J. 1995. Clinical recommendations using levels of evidence for antithrombotic agents. Chest 108 (4 Suppl): 227S-230S.

Crits-Christoph, P., Siqueland, L., Blaine, J., Frank, A., Luborsky, L., Onken, L.S., Muenz, L.R., Thase, M.E., Weiss, R.D., Gastfriend, D.R., Woody, G.E., Barber, J.P., Butler, S.F., Daley, D., Salloum, I., Bishop, S., Najavits, L.M., Lis, J., Mercer, D., Griffin, M.L., Moras, K. & Beck, A.T. 1999. Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse Collaborative Cocaine Treatment Study. Archives of General Psychiatry 56: 493-502.

Department of Health.1999. National Service Framework for Mental Health. London: HMSO.

Erwin, B.A., Heimberg, R.G., Juster, H.R. & Mindlin, M. 2002. Comorbid anxiety and mood disorders among persons with social anxiety disorder. Behaviour Research and Therapy 40: 19-35.

Fonagy, P. 1999. Process and outcome in mental health care delivery: A model approach to treatment evaluation. Bulletin of the Menninger Clinic 63: 288-304.

Gabbard, G.O., Gunderson, J.G. Fonagy, P. 2002. The place of psychoanalytic treatments within psychiatry. Archives of General Psychiatry 59: 505–510.

Gloaguen, V., Cottraux, J., Cucherat, M. & Blackburn, I.M. 1998. A meta-analysis of the effects of cognitive therapy in depressed patients. Journal of Affective Disorders 49: 59-72.

Guthrie, E. 2000. Psychotherapy for patients with complex disorders and chronic symptoms. The need for a new research paradigm. British Journal of Psychiatry 177: 131-137.

Guyatt, G.H., Sacket, D.L., Sinclair, J.C., Hayward, R., Cook, D.J. & Cook, R. 1995. User's guides to the medical literature. IX. A method for grading health care recommendations. Journal of the American Medical Association 274: 1800-1804.

Hahlweg, K., Fiegenbaum, W., Frank, M., Schroeder, B. & von Witzleben I. 2001. Short- and long-term effectiveness of an empirically supported treatment for agoraphobia. Journal of Consulting and Clinical Psychology 69: 375-82.

Hempel, C.G. 1970. On the "standard conception" of scientific theories. Minnesota Studies in the Philosophy of Science 4: 142-163.

Henry, W.P. 1998. Science, politics, and the politics of science: The use and misuse of empirically validated treatment research. Psychotherapy Research 8: 126-140.

Hilsenroth, M., Ackemman, SJ, Blagys, MD, Baity, M.R. & Mooney, M.A. 2003. Short-term psychodynamic psychotherapy for depression: An examination of statistical, clinically significant, and technique-specific change. Journal of Nervous and Mental Disease 1991: 349-357.

Hope, D.A., Herbert, J.D. & Withe, C. 1995. Diagnostic subtype, avoidant personality disorder, and efficacy of cognitive-behavioural group therapy for social phobia. Cognitive Therapy and Research 19: 399-417.

Hsu, L. 1989. Random sampling, randomisation, and equivalence of contrasted groups in psychotherapy outcome research. Journal of Consulting and Clinical Psychology 57: 131-137.

Jacobson, N.S. Dobson, K.S., Truax, P.A., Addis, M.E., Koerner, K., Gollan, J.K., Gortner, E. & Prince, S.E. 1996. A component analysis of cognitive-behavioural treatment for depression. Journal of Consulting and Clinical Psychology 64: 295-304.

Kendall, P.C., Holmbeck, G. & Verduin, T. 2004. Methodology, design, & evaluation in psychotherapy research. In Lambert, M.J. (Ed.). Bergin & Garfield´s Handbook of Psychotherapy & Behaviour Change, 5th ed. (pp. 16-43). New York, Wiley.

Kessler, R.C., Chiu, W.T., Demler, O. & Walters, E.E. 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV Disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry 62: 617-627.

Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C.B., Hughes, M., Eshleman, S., Wittchen, H.U. & Kendler, K.S. 1994. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Archives of General Psychiatry 51: 8-19.

Krupnick, J. L., Sotsky, S.M., Simmens, S., Moyer, J., Elkin, I., Watkins, J. & Pilkonis, P. 1996. The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. Journal of Consulting and Clinical Psychology 64: 532-539.

Lambert, M.J. & Bergin, A.E. 1994. The effectiveness of psychotherapy. In Bergin, A.E., & Garfield, S.L. (Eds.). Handbook of Psychotherapy & Behaviour Change, 4th ed. (pp. 143-189). New York, Wiley.

Leichsenring, F. 1985. Die probleme der externen validität in der psychotherapieforschung [The problems of external validity in psychotherapy research]. Zeitschrift für Klinische Psychologie 16: 214-227.

Leichsenring, F. 2004. Randomised controlled vs. effectiveness studies. A new research agenda. Bulletin of the Menninger Clinic 68: 137-151

Leichsenring, F. & Rabung, S. 2006. Change norms: A complementary approach to the issue of control groups in psychotherapy outcome research. Psychotherapy Research 16: 604-616.

Luborsky, L. 1984. Principles of Psychoanalytic Psychotherapy: A Manual of Supportive-Expressive Therapy. New York, Basic Books.

Lucock, M., Leach, C., Iveson, S., Lynch, K., Horsefield, C. & Hall, P. 2003. A systematic approach to practice-based evidence in a psychological therapies service. Clinical Psychology and Psychotherapy, 10, 389-399.

Nathan, P.E. & Gorman, J.M. (Eds). 2002. A Guide to Treatments that Work, 2nd ed. New York, Oxford University Press.

National Institute for Clinical Excellence (NICE). 2002. Clinical Guideline 1: Schizophrenia - Core Interventions in the Treatment and Management of Schizophrenia in Primary and Secondary Care (www.nice.org.uk).

Organista, K.C., Munoz, R.F. & Gonzales, G. 1994. Cognitive-behavioural therapy for depression in low-income and minority medical outpatients: Description of a program and exploratory analyses. Cognitive Therapy and Research 18: 241-259.

Orlinsky, H., Grawe, K. & Parks, B.K. 1994. Process and outcome in psychotherapy. In Bergin, A.E., & Garfield, S.L. (Eds.). Handbook of Psychotherapy & Behaviour change, 4th ed. (pp. 270-376). New York, Wiley.

Otto, M.W., Pollack, M.H., Gould, R.A., Worthington, J.J. 3rd, McArdle, E.T., Rosenbaum, J.F. & Heimberg, R.G. 2000. A comparison of the efficacy of clonazepam and cognitive-behavioural group therapy for the treatment of social phobia. Journal of Anxiety Disorders 14: 345-358.

Persons, J.B. & Silberschatz, G. 1998. Are results of randomised trials useful to psychotherapists? Journal of Consulting and Clinical Psychology 66: 126-135.

Persons, J.B., Bostrom. A. & Bertagnolli, A. 1999. Results of randomised controlled trials of cognitive therapy for depression generalize to private practice. Cognitive Therapy and Research 23: 535-548.

Peterson, A.L. & Halstead, T.S. 1998. Group cognitive behaviour therapy for depression in a community setting: A clinical replication series. Behaviour Therapy 29: 3-18.

Rosenthal, R. 1981. Pavlov's mice, Pfungst's horse, and Pygmalion's PONS: Some models for the study of interpersonal expectancy effects. In Sebeok, T.A., & Rosenthal, R. (Eds.). The Clever Hans Phenomenon: Communication with Horses, Whales, Apes and People. Annals of the New York Academy of Sciences 364: 182-198.

Roth, A.D. & Parry,G. 1997. The implications of psychotherapy research for clinical practice and service development: Lessons and limitations. Journal of Mental Health 6: 367-380.

Rothwell, P.M. 2005. External validity of randomised controlled trials. To whom do the results of this trial apply? Lancet 365: 82-92.

Schulz, K.F., Chalmers, I., Grimes, D.A. & Altman, D. 1994. Assessing the quality of randomisation from reports of controlled trials published in obstetrics and gynecology journals. Journal of the American Medical Association 13: 125-128.

Seligman, M.E.P. 1995. The effectiveness of psychotherapy. The Consumer Reports study. American Psychologist 50: 965-974.

Shadish, W.R., Cook, T.D. & Campbell, D.T. 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, Houghton Mifflin Company.

Shadish, W.R., Matt, G., Navarro, A. & Phillips, G. 2000. The effects of psychological therapies under clinically representative conditions: A meta-analysis. Journal of Consulting & Clinical Psychology 126: 512-529.

Sneed, J.D. 1971. The Logical Structure of Mathematical Physics. Dordrecht, Reidel.

Stegmüller, W. 1979. The Structuralist View of Theories. New York, Springer.

Stiles, W.B. & Shapiro, D.A. 1989. Abuse of the drug metaphor in psychotherapy process-outcome research. Clinical Psychology Review 9: 521-543.

Task Force on Promotion and Dissemination of Psychological Procedures. 1995. Training and dissemination of empirically-validated psychological treatments. Report and recommendations. Clinical Psychologist 48: 3-23.

Topolinski, S & Hertel, G. 2007. The Role of personality in psychotherapists´careers: Relationships between personality traits, therapeutic schools, and job satisfaction. Psychotherapy Research 17: 378-390.

Tuschen-Caffier, B., Pook, M. & Frank, M. 2001. Evaluation of manual-based cognitive-behavioural therapy for bulimia nervosa in a service setting. Behavioural Research & Therapy 39: 299-308.

Tynan, W.D., Schuman, W. & Lampert, N. 1999. Concurrent parent and child therapy groups for externalizing disorders: From the laboratory to the world of managed care. Cognitive and behavioural practice 6: 3-9.

U.K. Department of Health. 1996. NHS Psychotherapy Services in England: Review of Strategic Policy. London: Her Majesty´s Stationary Office.

Wade, W.A., Treat, T.A. & Stuart, G.L. 1998. Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. Journal of Consulting & Clinical Psychology 66: 231-239.

Wallerstein, R. 1999. Comment on Gunderson & Gabbard. Journal of the American Psychoanalytic Association 47: 728– 34.

Wells, K.B. 1999. Treatment research at the crossroads: The scientific interface of clinical trials and effectiveness research. American Journal of Psychiatry 156: 5-10.

Westen, D., Novotny, C.M. & Thompson-Brenner, H. 2004. The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. Psychological Bulletin 130: 631–63.

Westmeyer, H. 1982. Wissenschaftstheoretische aspekte der feldforschung [Scientific-theoretical aspects of field research]. In Patry, J.L. (Ed.). Feldforschung (pp. 67-84). Bern, Huber.

Westmeyer, H. 1989. Psychological theories from a structuralist point of view. In Westmeyer, H. (Ed.). Psychological Theories from a Structuralist Point of View (pp. 1-13). New York: Springer.